

Error analysis and minimum bound method for atmospheric remote sensing

Adrian Doicu*, Franz Schreier, Siegfried Hilgers, Michael Hess

DLR – German Aerospace Center, Remote Sensing Technology Institute, Oberpfaffenhofen, 82234 Weßling, Germany

Received 27 July 2004; received in revised form 22 July 2005; accepted 5 August 2005
Available online 8 June 2006

Abstract

In this study, we present an error analysis for Tikhonov regularization in a semi-stochastic setting. The analysis is carried out in such a way that it can be applied to any kind of inverse problem in atmospheric remote sensing. A method for selecting the optimal regularization parameter relying on the minimization of an estimator of the bound of the error between the first iterate and the exact solution is also discussed. Numerical simulations are performed for NO₂ retrieval from SCIAMACHY limb scatter measurements.

© 2006 Elsevier Ltd. All rights reserved.

Keywords: Inverse problems; Nonlinear least squares; Regularization; Remote sensing

1. Introduction

The optimal estimation method (otherwise known as the Bayesian approach) has a dominating role in atmospheric remote sensing (Rodgers, 2000). In this method, the errors of the solution are estimated in a stochastic setting and appear as covariance matrices.

When statistical information about atmospheric variability is poor, regularization methods accounting for deterministic information about the atmospheric state parameters can be used. One of the most efficient regularization methods for nonlinear ill-posed problems is Tikhonov regularization (Engl et al., 1996; Tautenhahn, 1994; Neubauer, 1989; O'Sullivan and Wahba, 1985). The selection of the regularization parameter is an important part of the method and a variety of a posteriori regularization parameter choice methods have been developed, e.g., the generalized cross validation method (O'Sullivan and Wahba, 1985), the discrepancy principle (Tautenhahn, 1997), the nonlinear L-curve criterion (Eriksson,

1996), and the minimum bound method. In the minimum bound method, the optimal regularization parameter is computed by minimizing a bound of the error of the solution. This method was originally formulated in a deterministic setting by Raus (1984) and in a semi-stochastic setting by Lukas (1998). A treatise of the method was presented in Engl et al. (1996). In all these studies the regularization matrix was the identity matrix. Unfortunately, the minimum bound method as an a posteriori regularization parameter choice rule is time consuming.

In Doicu et al. (2002) we used a version of Tikhonov regularization for solving inverse problems in atmospheric remote sensing, while in Doicu et al. (2003) we extended this approach to bound-constraint problems. Tikhonov regularization with B-splines for profile discretization has been presented in Doicu et al. (2005). In the present paper we present an error analysis in a semi-stochastic setting. Note that the error analysis in a deterministic or a semi-stochastic setting differs significantly from that in a complete stochastic setting. Further, we propose an a priori parameter choice method relying on the minimization of an estimator of the bound of the error between the first iterate and the exact solution. Similar approaches, but in a complete stochastic setting,

* Corresponding author. Tel.: +49 8153 28 3015; fax: +49 8153 28 1446.
E-mail address: adrian.doicu@dlr.de (A. Doicu).

has been discussed by, e.g., Steck (2002) and Carissimo et al. (2005).

The organization of our paper is as follows. Section 2 comprises the formulation of the inverse problem and the main concepts. In Section 3 we give a recipe for evaluating each component of the error, while in Section 4 the a priori regularization parameter choice method is presented. Numerical simulations concerning the retrieval of NO₂ profiles from limb scatter measurements are discussed in Section 5, the application to real measurement data is presented in Section 6, and Section 7 summarizes our results. Mathematical details are treated in the appendices. Vectors and matrices are indicated by lower and upper case boldface letters, respectively.

2. Inverse problem

A common retrieval problem in atmospheric remote sensing is to estimate vertical profiles of atmospheric parameters from spectroscopic measurements. From a computational point of view the basic problem is the inversion of observed spectra to obtain profiles of constituent concentrations. The discretization of the radiative transfer equation leads to the data model:

$$\mathbf{y} = \mathbf{F}(\mathbf{x}) \quad (1)$$

where the mapping $\mathbf{F} : \mathbb{R}^n \rightarrow \mathbb{R}^m$ represents the forward model, $\mathbf{y} \in \mathbb{R}^m$ is the exact data vector, $\mathbf{x} \in \mathbb{R}^n$ is the state vector containing the discrete representation of the atmospheric profile to be retrieved (e.g., molecular density profiles at different altitudes), and \mathbb{R}^n is the n -dimensional real Euclidean space. The exact data are assumed to be attainable, i.e., there exists the exact solution $\hat{\mathbf{x}}$ such that $\mathbf{y} = \mathbf{F}(\hat{\mathbf{x}})$. Measurements are made to a finite accuracy and in practice only the noisy data vector

$$\mathbf{y}^\delta = \mathbf{y} + \boldsymbol{\delta}, \quad (2)$$

is available. In our analysis we consider a semi-stochastic data model in the sense that the exact solution $\hat{\mathbf{x}}$ is deterministic but the measurement error $\boldsymbol{\delta}$ is stochastic with zero mean and the covariance matrix $\mathbf{S}_\delta = \mathcal{E}\{\boldsymbol{\delta} \cdot \boldsymbol{\delta}^T\} = (1/m)\mathbf{I}_m$, where \mathcal{E} is the expected value operator and \mathbf{I}_m is the identity matrix (of rank m). In general, if the measurement error is described by a symmetric and positive definite covariance matrix \mathbf{S}_δ , one can obtain a “normalized” data model with identity covariance matrix by using the prewhitening technique (Rodgers, 2000).

The inverse ill-posed problem is solved in the least squares sense by means of Tikhonov regularization. In this approach an approximate solution $\mathbf{x}_\lambda^\delta$ is computed by minimizing the function

$$\mathcal{F}(\mathbf{x}) = \frac{1}{2} \left[\|\mathbf{F}(\mathbf{x}) - \mathbf{y}^\delta\|^2 + \lambda^2 \|\mathbf{L}(\mathbf{x} - \mathbf{x}_a)\|^2 \right], \quad (3)$$

where \mathbf{L} is the regularization matrix, λ is the regularization parameter, and \mathbf{x}_a is the a priori state vector, the best beforehand

estimator of $\hat{\mathbf{x}}$. The Gauss–Newton method applied to the above minimization problem leads to the iterative process

$$\begin{aligned} \mathbf{x}_{k+1\lambda}^\delta &= \mathbf{x}_{k\lambda}^\delta - [\mathbf{K}^T(\mathbf{x}_{k\lambda}^\delta)\mathbf{K}(\mathbf{x}_{k\lambda}^\delta) + \lambda^2\mathbf{L}^T\mathbf{L}]^{-1} \\ &\quad \times [\mathbf{K}^T(\mathbf{x}_{k\lambda}^\delta)(\mathbf{F}(\mathbf{x}_{k\lambda}^\delta) - \mathbf{y}^\delta) + \lambda^2\mathbf{L}^T\mathbf{L}(\mathbf{x}_{k\lambda}^\delta - \mathbf{x}_a)], \end{aligned} \quad (4)$$

where $\mathbf{K}(\mathbf{x}) = \mathbf{F}'(\mathbf{x})$ is the Jacobian matrix evaluated at \mathbf{x} . Essentially, at each iteration step we consider a linear problem, and $\mathbf{x}_{k+1\lambda}^\delta$ minimizes the function

$$\mathcal{F}_k^l(\mathbf{x}) = \|\mathbf{F}(\mathbf{x}_{k\lambda}^\delta) - \mathbf{y}^\delta + \mathbf{K}(\mathbf{x}_{k\lambda}^\delta)(\mathbf{x} - \mathbf{x}_{k\lambda}^\delta)\|^2 + \lambda^2 \|\mathbf{L}(\mathbf{x} - \mathbf{x}_a)\|^2. \quad (5)$$

3. Error analysis

The accuracy of a retrieval method can be characterized by the discrepancy between the approximate solution $\mathbf{x}_\lambda^\delta$ and the exact solution $\hat{\mathbf{x}}$. In this section we derive estimates of the error components in a semi-stochastic setting and for a general regularization matrix \mathbf{L} . The only assumption on \mathbf{L} is that the matrix $\mathbf{L}^T\mathbf{L}$ possesses an inverse. Assuming that the sequence of iterates $(\mathbf{x}_{k\lambda}^\delta)$ converges toward $\mathbf{x}_\lambda^\delta$, i.e., $\mathbf{x}_{k\lambda}^\delta \rightarrow \mathbf{x}_\lambda^\delta$ as $k \rightarrow \infty$, we see that $\mathbf{x}_\lambda^\delta$ satisfies the first order optimality condition

$$\mathbf{K}^T(\mathbf{x}_\lambda^\delta) [\mathbf{F}(\mathbf{x}_\lambda^\delta) - \mathbf{y}^\delta] + \lambda^2 \mathbf{L}^T \mathbf{L} (\mathbf{x}_\lambda^\delta - \mathbf{x}_a) = 0. \quad (6)$$

The data model can be linearized around $\mathbf{x}_\lambda^\delta$, i.e.,

$$\mathbf{y} = \mathbf{F}(\mathbf{x}_\lambda^\delta) + \mathbf{K}(\mathbf{x}_\lambda^\delta)(\hat{\mathbf{x}} - \mathbf{x}_\lambda^\delta) + \mathcal{R}(\mathbf{x}_\lambda^\delta, \hat{\mathbf{x}}), \quad (7)$$

with $\mathcal{R}(\mathbf{x}_\lambda^\delta, \hat{\mathbf{x}})$ being the linearization error. Taking into account Eqs. (2) and (6), and neglecting the linearization error in Eq. (7) yields

$$\mathbf{x}_\lambda^\delta - \hat{\mathbf{x}} = (\mathbf{A}_\lambda - \mathbf{I}_n)(\hat{\mathbf{x}} - \mathbf{x}_a) + \mathbf{K}_\lambda^\dagger \boldsymbol{\delta}, \quad (8)$$

where $\mathbf{K}_\lambda^\dagger$ is the regularized generalized inverse or the gain matrix,

$$\mathbf{K}_\lambda^\dagger = (\mathbf{K}^T \mathbf{K} + \lambda^2 \mathbf{L}^T \mathbf{L})^{-1} \mathbf{K}^T, \quad (9)$$

and \mathbf{A}_λ is the averaging kernel,

$$\mathbf{A}_\lambda = \mathbf{K}_\lambda^\dagger \mathbf{K} = (\mathbf{K}^T \mathbf{K} + \lambda^2 \mathbf{L}^T \mathbf{L})^{-1} \mathbf{K}^T \mathbf{K}. \quad (10)$$

In order to simplify the notations we omit to indicate the dependency of the Jacobian matrix on the evaluation point that will be clear from the context. The expression of the error in $\hat{\mathbf{x}}$ can be derived by rewriting Eq. (8) as

$$\boldsymbol{\epsilon}_{\text{total}}^\delta \equiv \mathbf{x}_\lambda^\delta - \hat{\mathbf{x}} = (\mathbf{A}_\lambda - \mathbf{I}_n)(\hat{\mathbf{x}} - \mathbf{x}_a) + \mathbf{K}_\lambda^\dagger \boldsymbol{\delta} = \boldsymbol{\epsilon}_{\text{smooth}} + \boldsymbol{\epsilon}_{\text{noise}}^\delta. \quad (11)$$

Assuming that the Jacobian matrix evaluated at $\mathbf{x}_\lambda^\delta$ does not depend on $\boldsymbol{\delta}$, i.e., $\mathbf{K}(\mathbf{x}_\lambda^\delta) \approx \mathbf{K}(\hat{\mathbf{x}})$, we see that the smoothing error $\boldsymbol{\epsilon}_{\text{smooth}}$ is a deterministic quantity. The total error $\boldsymbol{\epsilon}_{\text{total}}^\delta$ is

stochastic with the mean $\mathcal{E}\{\boldsymbol{\epsilon}_{\text{total}}^\delta\} = \boldsymbol{\epsilon}_{\text{smooth}}$ and the covariance matrix $\mathbf{S}(\boldsymbol{\epsilon}_{\text{total}}^\delta) = \mathbf{S}(\boldsymbol{\epsilon}_{\text{noise}}^\delta)$. Upon squaring Eq. (11) it is apparent that

$$\|\boldsymbol{\epsilon}_{\text{total}}^\delta\|^2 \leq 2(\|\boldsymbol{\epsilon}_{\text{smooth}}\|^2 + \|\boldsymbol{\epsilon}_{\text{noise}}^\delta\|^2). \quad (12)$$

Applying the expected value operator yields

$$\mathcal{E}\{\|\boldsymbol{\epsilon}_{\text{total}}^\delta\|^2\} \leq 2(\|\boldsymbol{\epsilon}_{\text{smooth}}\|^2 + \mathcal{E}\{\|\boldsymbol{\epsilon}_{\text{noise}}^\delta\|^2\}) \quad (13)$$

and a bound of the total error can be defined as

$$Y = 2(\|\boldsymbol{\epsilon}_{\text{smooth}}\|^2 + \mathcal{E}\{\|\boldsymbol{\epsilon}_{\text{noise}}^\delta\|^2\}). \quad (14)$$

In Appendix A we show that an estimator of the smoothing error vector is

$$\boldsymbol{\epsilon}_{\text{smooth}}^\delta = \mathbf{K}^\dagger \mathbf{r}_\lambda^\delta, \quad (15)$$

where $\mathbf{K}^\dagger = (\mathbf{K}^T \mathbf{K})^{-1} \mathbf{K}^T$ is the generalized inverse and $\mathbf{r}_\lambda^\delta = \mathbf{F}(\mathbf{x}_\lambda^\delta) - \mathbf{y}^\delta$ is the residual error at $\mathbf{x}_\lambda^\delta$. Since $\boldsymbol{\epsilon}_{\text{smooth}}^\delta$ is an estimator of $\boldsymbol{\epsilon}_{\text{smooth}}$ depending on the measurement error $\boldsymbol{\delta}$, Y^δ will be an estimator of the bound of the total error Y . Eq. (15) shows that the norm of the smoothing error is bounded by the norm of the residual error. As shown in Appendix A, $\boldsymbol{\epsilon}_{\text{smooth}}^\delta$ does not converge to zero as λ tends to zero. However, the residual error norm $\|\mathbf{r}_\lambda^\delta\|$ increases with increasing λ (Hansen, 1998), and therefore we may conclude that the smoothing error norm $\|\boldsymbol{\epsilon}_{\text{smooth}}^\delta\|$ is an increasing function of the regularization parameter λ .

The derivation of the noise error is standard (e.g. Rodgers, 2000). Using $\mathcal{E}\{\boldsymbol{\delta} \cdot \boldsymbol{\delta}^T\} = (1/m)\mathbf{I}_m$, we obtain

$$\mathcal{E}\{\|\boldsymbol{\epsilon}_{\text{noise}}^\delta\|^2\} = \frac{1}{m} \text{trace}\{\mathbf{K}_\lambda^{\dagger T} \mathbf{K}_\lambda^\dagger\} = \frac{1}{m} \text{trace}\{\mathbf{K}_\lambda^\dagger \mathbf{K}_\lambda^{\dagger T}\}, \quad (16)$$

and

$$\mathbf{S}(\boldsymbol{\epsilon}_{\text{noise}}^\delta) = (1/m)\mathbf{K}_\lambda^\dagger \mathbf{K}_\lambda^{\dagger T}. \quad (17)$$

In the specific case $\mathbf{L} = \mathbf{I}_n$, we have

$$\mathcal{E}\{\|\boldsymbol{\epsilon}_{\text{noise}}^\delta\|^2\} \leq (C/\lambda^2),$$

where $C \leq 1$ (Appendix B). This estimate indicates that the noise error is a decreasing function of the regularization parameter λ .

A numerical robust method for error estimation relies on the use of the generalized singular value decomposition (GSVD) of the matrix pair (\mathbf{K}, \mathbf{L}) (Hansen, 1998). We recall that if $\mathbf{K} \in \mathbb{R}^{m \times n}$, $\mathbf{L} \in \mathbb{R}^{p \times n}$ and $m \geq n \geq p$, the GSVD of the matrix pair (\mathbf{K}, \mathbf{L}) is given by

$$\mathbf{K} = \mathbf{U} \boldsymbol{\Sigma}_1 \mathbf{Z}^{-1} \text{ and } \mathbf{L} = \mathbf{V} \boldsymbol{\Sigma}_2 \mathbf{Z}^{-1}, \quad (18)$$

where $\mathbf{U} \in \mathbb{R}^{m \times m}$ and $\mathbf{V} \in \mathbb{R}^{p \times p}$ are unitary matrices, $\mathbf{Z} \in \mathbb{R}^{n \times n}$ is a non-singular matrix, $\boldsymbol{\Sigma}_1 \in \mathbb{R}^{m \times n}$ and $\boldsymbol{\Sigma}_2 \in \mathbb{R}^{p \times n}$ are diagonal matrices

$$\boldsymbol{\Sigma}_1 = \begin{bmatrix} \mathbf{I}_{n-p} & 0 \\ 0 & \text{diag}(\alpha_i) \\ 0 & 0 \end{bmatrix}, \quad \boldsymbol{\Sigma}_2 = [0 \quad \text{diag}(\beta_i)],$$

$\text{diag}(\alpha_i), \text{diag}(\beta_i) \in \mathbb{R}^{p \times p}$ and the ratios $\gamma_i = \alpha_i/\beta_i$, $i = n-p+1, \dots, n$, are the generalized singular values. Then we have

$$\begin{aligned} \boldsymbol{\epsilon}_{\text{smooth}}^\delta &= \sum_{i=1}^{n-p} \langle \mathbf{u}_i, \mathbf{r}_\lambda^\delta \rangle \mathbf{z}_i + \sum_{i=n-p+1}^n \frac{1}{\alpha_i} \langle \mathbf{u}_i, \mathbf{r}_\lambda^\delta \rangle \mathbf{z}_i, \\ \mathcal{E}\{\|\boldsymbol{\epsilon}_{\text{noise}}^\delta\|^2\} &= \frac{1}{m} \sum_{i=1}^n \sum_{j=1}^n (\kappa_j z_{ij})^2, \\ \mathbf{S}(\boldsymbol{\epsilon}_{\text{noise}}^\delta) &= (1/m)\mathbf{Z} \boldsymbol{\kappa} \mathbf{Z}^T. \end{aligned} \quad (19)$$

where \mathbf{u}_j and \mathbf{z}_j are the columns of \mathbf{U} and \mathbf{Z} , respectively, z_{ij} are the entries of the matrix \mathbf{Z} , $\boldsymbol{\kappa} = \text{diag}(\kappa_j)$, and

$$\kappa_j = \begin{cases} 1 & j = 1, 2, \dots, n-p \\ \frac{1}{\alpha_j} \frac{\gamma_j^2}{\gamma_j^2 + \lambda^2} & \text{for } j = n-p+1, \dots, n. \end{cases} \quad (20)$$

3.1. Comments

- (1) The quantities $\|\boldsymbol{\epsilon}_{\text{smooth}}^\delta\|$ and $\sqrt{\mathcal{E}\{\|\boldsymbol{\epsilon}_{\text{noise}}^\delta\|^2\}}$ can be used as a posteriori global estimators of the components of the total error, while $\boldsymbol{\epsilon}_{\text{smooth}}^\delta$ and the square root of the diagonal of the noise covariance matrix $\mathbf{S}(\boldsymbol{\epsilon}_{\text{noise}}^\delta)$ can be used as a posteriori altitude-dependent estimators of the smoothing and noise errors, respectively.
- (2) Since the smoothing error increases with increasing λ and the noise error decreases with increasing λ , the total error has a minimum for an optimal value of λ . This value can be determined by minimizing the estimator of the bound of the total error with respect to the regularization parameter λ . The minimization condition leads to a nonlinear equation in λ (see Eq. (15) in Appendix B for the case $\mathbf{L} = \mathbf{I}_n$). This a posteriori regularization parameter choice method is known as the minimum bound method. The numerical realization of Tikhonov regularization with λ chosen as the solution of the minimum bound equation requires to solve the nonlinear minimization problem (Eq. (3)) several times for different regularization parameters λ . This strategy is time consuming and not appropriate for the operational usage of a retrieval processor.
- (3) In our analysis, we used $\boldsymbol{\epsilon}_{\text{smooth}}^\delta$ as an estimator for $\boldsymbol{\epsilon}_{\text{smooth}}$ because $\boldsymbol{\epsilon}_{\text{smooth}}^\delta$ is a computable quantity. Note that the same problem appears in a complete stochastic setting. In this case

$$\mathcal{E}\{\|\boldsymbol{\epsilon}_{\text{smooth}}^\delta\|^2\} = \text{trace}\{\mathbf{S}_{\text{smooth}}\},$$

$$\mathbf{S}_{\text{smooth}} = (\mathbf{A}_\lambda - \mathbf{I}_n) \mathbf{S}_e (\mathbf{A}_\lambda - \mathbf{I}_n)^T,$$

where $\mathbf{S}_{\text{smooth}}$ is the covariance matrix of the smoothing error and \mathbf{S}_e is the covariance of the ensemble of states about the

mean state. Often the true covariance matrix S_e is not known and the estimation of the smoothing error is inaccurate.

4. A priori regularization parameter choice method

A regularization parameter choice method that avoids the multiple solution of the nonlinear minimization problem (Eq. (3)) consists in the determination of the optimal value of λ from an estimator of the bound of the error between the first iterate and the exact solution. By convention, we refer to the error between the first iterate and the exact solution as the a priori error. To derive an estimator of the bound of the a priori error we consider the Gauss–Newton iteration (Eq. (4)) and assume that the initial guess coincides with the a priori state vector, that is, we set $\mathbf{x}_0 = \mathbf{x}_a$. The first iterate is given by

$$\mathbf{x}_{1\lambda}^{\delta} - \mathbf{x}_a = [\mathbf{K}^T(\mathbf{x}_a)\mathbf{K}(\mathbf{x}_a) + \lambda^2\mathbf{L}^T\mathbf{L}]^{-1}\mathbf{K}^T(\mathbf{x}_a)(\mathbf{y}^{\delta} - \mathbf{F}(\mathbf{x}_a)). \quad (21)$$

Further, linearizing the data model around the a priori state yields

$$\mathbf{y} = \mathbf{F}(\mathbf{x}_a) + \mathbf{K}(\mathbf{x}_a)(\hat{\mathbf{x}} - \mathbf{x}_a) + \mathcal{R}(\mathbf{x}_a, \hat{\mathbf{x}}). \quad (22)$$

Neglecting $\mathcal{R}(\mathbf{x}_a, \hat{\mathbf{x}})$ and using $\mathbf{y}^{\delta} = \mathbf{y} + \boldsymbol{\delta}$, gives

$$\mathbf{x}_{1\lambda}^{\delta} - \hat{\mathbf{x}} = (\mathbf{A}_{\lambda} - \mathbf{I}_n)(\hat{\mathbf{x}} - \mathbf{x}_a) + \mathbf{K}_{\lambda}^{\dagger}\boldsymbol{\delta}. \quad (23)$$

This relation is similar to Eq. (8). The major difference is that the Jacobian matrix is computed at the a priori state. The expected value of the a priori noise error is given by Eqs. (9) and (16) with $\mathbf{K} = \mathbf{K}(\mathbf{x}_a)$. Because we do not use an optimality condition, the expression of the a priori smoothing error differs from that given by Eq. (15). To compute the a priori smoothing error we use Eq. (22) and express $\epsilon_{\text{smooth}}^a$ as

$$\begin{aligned} \epsilon_{\text{smooth}}^a &= (\mathbf{A}_{\lambda} - \mathbf{I}_n)(\hat{\mathbf{x}} - \mathbf{x}_a) \\ &= -(\mathbf{A}_{\lambda} - \mathbf{I}_n)(\mathbf{K}^T\mathbf{K})^{-1}\mathbf{K}^T[\mathbf{F}(\mathbf{x}_a) - \mathbf{y}]. \end{aligned} \quad (24)$$

Replacing \mathbf{y} by \mathbf{y}^{δ} , we obtain after straightforward calculations the following practical estimator

$$\epsilon_{\text{smooth}}^{a\delta} = (\mathbf{K}^{\dagger} - \mathbf{K}_{\lambda}^{\dagger})\mathbf{r}_a^{\delta}, \quad (25)$$

where $\mathbf{r}_a^{\delta} = \mathbf{F}(\mathbf{x}_a) - \mathbf{y}^{\delta}$ is the residual error at \mathbf{x}_a . In contrast to the estimator $\epsilon_{\text{smooth}}^{\delta}$, $\epsilon_{\text{smooth}}^{a\delta}$ converges to zero as λ tends to zero. In terms of the GSVD of the matrix pair (\mathbf{K}, \mathbf{L}) , we have

$$\epsilon_{\text{smooth}}^{a\delta} = \sum_{i=n-p+1}^n \frac{1}{\alpha_i} \frac{\lambda^2}{\gamma_i^2 + \lambda^2} \langle \mathbf{u}_i, \mathbf{r}_a^{\delta} \rangle \mathbf{z}_i. \quad (26)$$

The a priori smoothing error is also an increasing function of the regularization parameter λ . Here, the residual error \mathbf{r}_a^{δ} does not depend on λ , but the filter function $f(\lambda) = \lambda^2 / (\gamma_i^2 + \lambda^2)$ is an increasing function of λ for all γ_i with $i = n - p + 1, \dots, n$. Consequently, the optimal regularization parameter can be computed at the first iteration by minimizing the estimator of the bound of the a priori error (Eq. (14)) with the a priori smoothing error estimator given by Eq. (26). This parameter

choice method will be referred to as the a priori minimum bound method.

The derivation of the a priori error has been carried out under a strong linearity assumption. Neglecting the linearization error in Eq. (22) means that the forward model is assumed to be linear around the a priori state. Note that this assumption is used for selecting the regularization parameter, but not for finding a solution.

We expect this a priori parameter choice method to be valid only for problems with moderate degrees of nonlinearity. In a stochastic setting, the degree of nonlinearity for any particular problem can be examined by comparing the forward model with the linearized forward model within either the a priori variability or the solution error covariance (Rodgers, 2000). In a semi-stochastic setting the degree of nonlinearity can be explored a posteriori by comparing the linearization error $\mathcal{R}(\mathbf{x}_a, \mathbf{x})$ in a ball of radius $\rho = |\mathbf{x}_{\lambda}^{\delta} - \mathbf{x}_a|$ with S_{δ} . This can be done systematically by evaluating the cost function $\chi^2 = \mathcal{R}^T S_{\delta} \mathcal{R}$. However, in our analysis we will use a more pragmatic criterion. By convention, the degree of nonlinearity of a specific problem is given by the number of iterations required to achieve convergence. A problem has a small degree of nonlinearity if the number of iterations does not exceed 2, and a moderate degree of nonlinearity if the number of iterations is 3 or 4.

The strategy for selecting the regularization parameter at the first iteration step is in agreement with the technique described in Doicu et al. (2002). There, we derived the optimal regularization parameter by using the L-curve method for the linear subproblem given by Eq. (5). The L-curve method consists in the analysis of the graph of the curve obtained by plotting the smoothing norm versus the residual norm (Hansen, 1998). This curve exhibits a typical ‘‘L’’ shape, and the optimal value of the regularization parameter corresponds to the corner of the ‘‘L’’ (the point with maximum curvature). For linear problems, the regularization parameters corresponding to the L-curve’s corner are close to the regularization parameter balancing the smoothing and noise errors (Hansen, 1998). Therefore, for linear problems, the a priori minimum bound method and the L-curve method will give similar results.

5. Numerical simulation

The forward problem of atmospheric remote sensing essentially describes the radiative transfer in the atmosphere. The algorithms developed in Doicu et al. (2002, 2003, 2005) have been applied to inverse problems in infrared atmospheric sounding, where the radiative transfer is dominated by molecular absorption (Schreier and Böttger, 2003; Amato et al., 2003). Here we consider atmospheric retrievals in the ultraviolet and visible spectral domains, where multiple scattering is the dominant mechanism of radiative transfer (cf., e.g., Spurr et al., 2000).

Several satellite instruments measure ultraviolet and visible solar radiation scattered from the Earth’s limb to monitor vertical profiles of O_3 and other minor constituents such as NO_2 .

A sequence of observations (limb scan) corresponding to different tangent altitudes are used for atmospheric retrieval. One instrument that employs this technique is the SCanning Imaging Absorption SpectroMeter for Atmospheric Chartography (SCIAMACHY) aboard the ENVISAT satellite (Bovensmann et al., 1999). The radiative transfer model for SCIAMACHY limb radiance simulation is a single-scattering model, while the multiple scattering effect is taken into account by using look-up table corrections (Flittner et al., 2000). The Jacobians required for the iterative solution of the nonlinear least squares problem are computed analytically; for an alternative approach based on perturbation theory we refer to Box (2002). The data model Eq. (1) is expressed in terms of normalized limb radiances and relies on the DOAS (Differential Optical Absorption Spectroscopy) technique (Platt, 1994) i.e.,

$$\mathbf{y}^\delta = \ln(\mathbf{I}_{\text{meas}}^n) - \mathbf{P}_{\text{meas}},$$

$$\mathbf{F} = \ln(\mathbf{I}_{\text{sim}}^n) - \mathbf{P}_{\text{sim}}.$$

The normalized vector of limb radiances is given by $\mathbf{I}^n = \mathbf{I}/\mathbf{I}^{\text{ref}}$ where the division is component-wise, $\mathbf{I}^{\text{ref}} = \mathbf{I}^{\text{ref}}(\lambda)$ is the reference radiance spectrum for all wavelengths λ of interest and $\mathbf{I} = \mathbf{I}(s, \lambda)$ is the vector of radiance spectra for all tangent heights s except for the reference tangent height s_{ref} . \mathbf{P} is a lower-order polynomial in λ , whose coefficients are obtained by fitting the function $\ln(\mathbf{I}^n)$ for a specific tangent height in the wavelength domain. Note that the spectra normalization eliminates solar Fraunhofer structure, the instrument response function and the need for an absolute instrument calibration. By subtracting lower-order polynomials from radiance spectra we remove those spectral features that are smoothly varying functions of wavelength, particularly those due to Rayleigh and Mie scatterings. The DOAS type model is very sensitive to weak absorptions and has a moderate degree of nonlinearity (the number of iterations required to achieve convergence does not exceed 4).

In this section we consider numerical experiments concerning the retrieval of NO_2 profile from simulated SCIAMACHY data in the visible domain. Specifically, 168 equidistant points are considered in the 425–460 nm range within SCIAMACHY Channel 3. The noisy data vector (i.e., the contaminated spectrum) is generated by using the noise variances corresponding to this channel of the SCIAMACHY instrument. The a priori and initial gas profile were assumed to be identical and were chosen from the U.S. standard atmosphere. The exact gas profiles $\hat{\mathbf{x}}$ were taken as a scaled version of the a priori profiles by a factor of 1.2, i.e. $\hat{\mathbf{x}} = 1.2\mathbf{x}_a$. This choice leads to a problem with a small degree of nonlinearity since 2 iterations are required to achieve convergence. An equidistant altitude retrieval grid with $n = 14$ grid points between 15.75 and 61.25 km is considered. The number of limb spectra is 14 and the corresponding tangent heights vary between 16 and 61.5 km in steps of 3.5 km.

To estimate the accuracy of a computed solution $\mathbf{x}_\lambda^\delta$ with respect to the exact solution $\hat{\mathbf{x}}$, we use the relative error

$$\mathcal{A}_x = \frac{\|\mathbf{x}_\lambda^\delta - \hat{\mathbf{x}}\|}{\|\hat{\mathbf{x}}\|}.$$

Similarly, we define

$$\mathcal{A}_{\text{smooth}} = \frac{\|\boldsymbol{\epsilon}_{\text{smooth}}^\delta\|}{\|\hat{\mathbf{x}}\|},$$

$$\mathcal{A}_{\text{noise}} = \frac{\sqrt{\mathcal{E}\{\|\boldsymbol{\epsilon}_{\text{noise}}^\delta\|^2\}}}{\|\hat{\mathbf{x}}\|},$$

$$\mathcal{A}_{\text{total}} = \frac{\sqrt{\mathbf{Y}^\delta}}{\|\hat{\mathbf{x}}\|}$$

as the relative smoothing error, relative noise error and relative bound of the total error, respectively. The relative a priori errors are defined in an analogous manner and are marked with the superscript a.

In Fig. 1 we plot the relative a priori errors $\mathcal{A}_{\text{smooth}}^a$ and $\mathcal{A}_{\text{noise}}^a$ for different values of the regularization parameter λ . It can be seen that the relative a priori noise error decreases with increasing λ , while the relative a priori smoothing error increases (almost) with increasing λ . The relative bound of the a priori total error $\mathcal{A}_{\text{total}}^a$ has a minimum value for $\lambda_{\text{opt}}^a = 0.227$. For this value of the regularization parameter we compute the approximate solution $\mathbf{x}_{\lambda_{\text{opt}}^a}^\delta$ and obtain a relative error with respect to the exact solution of $\mathcal{A}_x = 1.47\%$. The retrieved profiles are shown in Fig. 2. In Fig. 3 we plot the L-curve at the first iteration, i.e., the graph of smoothing norm versus the residual norm. The L-curve has a pronounced corner and the optimal value of the regularization parameter λ_{opt}^a is near to the corner of “L”. Thus, for this example, the a priori minimum bound method is equivalent to the L-curve method.

In Fig. 4 we plot the relative errors $\mathcal{A}_{\text{smooth}}$ and $\mathcal{A}_{\text{noise}}$ of the solutions obtained by solving the nonlinear minimization problem several times for different regularization parameters

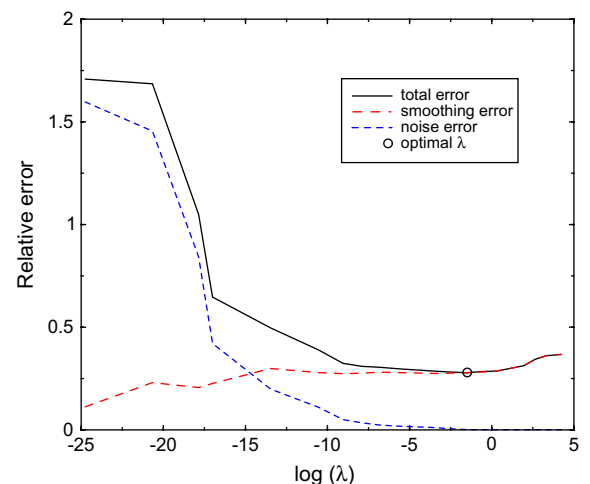


Fig. 1. Relative a priori errors (smoothing error, noise error and the estimator of the bound of the total error) for different values of the regularization parameter.

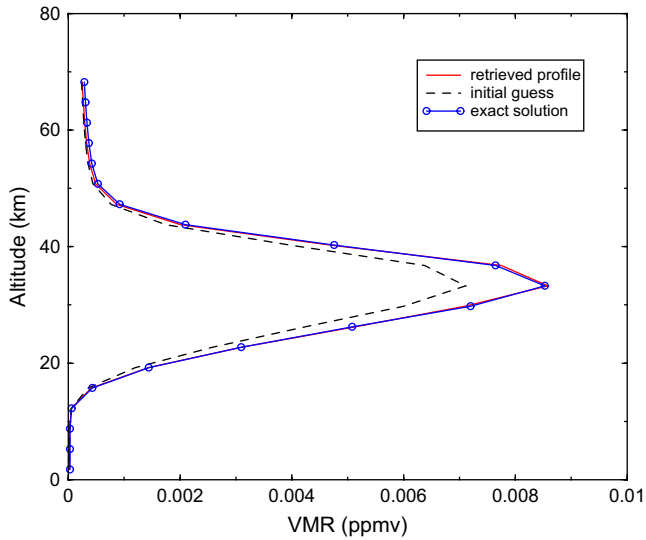


Fig. 2. Retrieved NO₂ profile for the optimal regularization parameter computed with the a priori minimum bound method.

λ . As before, the relative bound of the total error Δ_{total} possesses a minimum that is attained for $\lambda_{\text{opt}} = 0.178$. In this case, the relative error with respect to the exact solution is $\Delta_x = 1.64\%$. Evidently, the a priori regularization parameter choice method estimates a regularization parameter which is very close to the minimizer of the bound of the total error.

In Fig. 5 we plot the relative error Δ_x for the above sequence of regularization parameters. The minimum of this error is $\Delta_x = 1.38\%$ and corresponds to $\lambda = 0.301$. It was surprising to find that the optimal regularization parameter estimated by minimizing the bound of the a priori total error is more close to the optimal value than the minimizer of the bound of the total error. This suggests that $\epsilon_{\text{smooth}}^{\text{a}\delta}$ is a better estimator for $\epsilon_{\text{smooth}}^{\text{a}}$ than $\epsilon_{\text{smooth}}^{\delta}$ for ϵ_{smooth} . However, all solutions are of comparable accuracy and the differences between them are generally small.

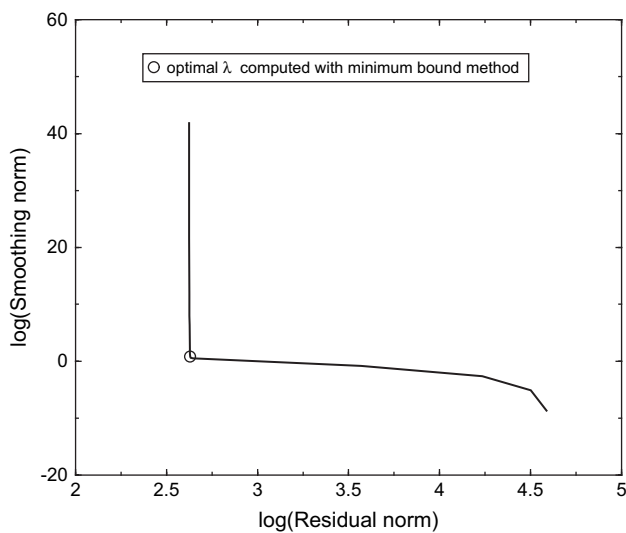


Fig. 3. L-curve and the optimal regularization parameter computed with the a priori minimum bound method.

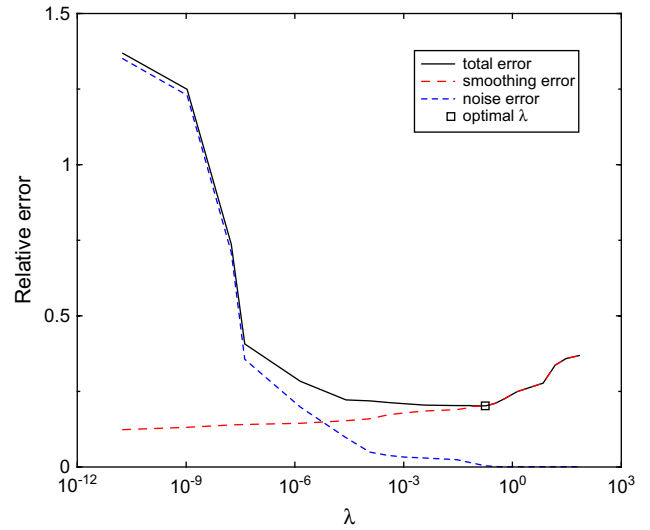


Fig. 4. Relative a posteriori errors (smoothing error, noise error and the estimator of the bound of the total error) for different values of the regularization parameter.

6. SCIAMACHY measurements

In this section SCIAMACHY measurements for the orbit 12253 on 4 July 2004 are considered. The goal of our analysis is to compare the performances of Tikhonov regularization with the a priori minimum bound method and the L-curve method for real data. Note that for data with significant noise level, r_λ^δ is not an accurate estimator of r_λ and the smoothing error is potentially overestimated. The geographical coordinates of the exemplary measurements (“states”) considered in our analysis are given in Table 1. For these simulations, the a priori information is obtained from the dataset of McLinden et al. (2002). The number of limb spectra is 9, while the tangent altitudes are state dependent and vary

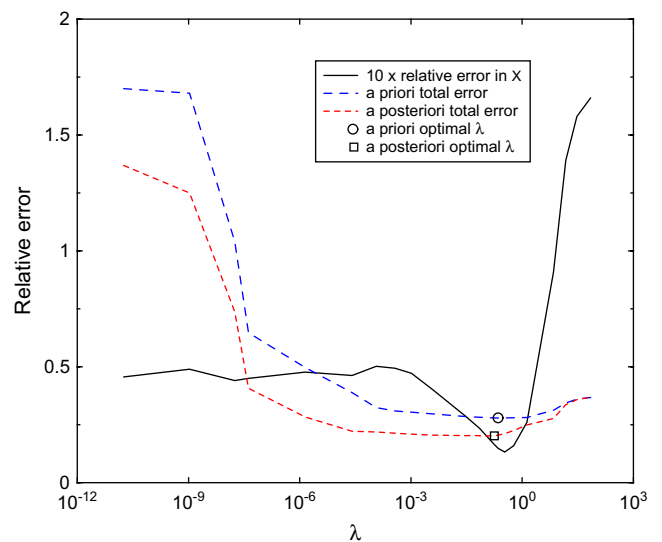


Fig. 5. Relative error with respect to the exact solution and the bound estimators of the a priori and a posteriori total error. Note that the relative error Δ_x has been scaled by a factor of 10 for better illustration of its minimum.

Table 1
Geographical coordinates of SCIAMACHY observations

State	Latitude	Longitude
	deg N	deg E
1	79	145
2	67	112
3	53	102
4	39	96
5	24	92
6	9	88

between 20 and 47 km. Table 2 illustrates as an example, the tangent heights for the measurement at the latitude of 79 deg N. The altitude retrieval grid is equidistant and consists of $n = 9$ grid points between 19.25 and 47.25 km. Below and above the retrieval limits, an a priori profile is considered.

Denoting by $\mathbf{x}_{\lambda\text{MB}}^\delta$ and $\mathbf{x}_{\lambda\text{LC}}^\delta$ the minimum bound and the L-curve solutions, we use the relative error vector

$$[\mathbf{A}_x]_i = \frac{|[\mathbf{x}_{\lambda\text{MB}}^\delta]_i - [\mathbf{x}_{\lambda\text{LC}}^\delta]_i|}{[\mathbf{x}_{\lambda\text{MB}}^\delta]_i},$$

the relative smoothing and noise error vectors

$$[\mathbf{A}_{\text{smooth}}]_i = \frac{|[\boldsymbol{\epsilon}_{\text{smooth}}^\delta]_i|}{[\mathbf{x}_{\lambda}^\delta]_i},$$

$$[\mathbf{A}_{\text{noise}}]_i = \frac{\sqrt{[\mathbf{S}(\boldsymbol{\epsilon}_{\text{noise}}^\delta)]_{ii}}}{[\mathbf{x}_{\lambda}^\delta]_i},$$

with $i = 1, 2, \dots, n$, and the bound of the total error vector

$$\mathbf{A}_{\text{total}} = \mathbf{A}_{\text{smooth}} + \mathbf{A}_{\text{noise}},$$

to characterize the solutions. The retrieved profiles together with the altitude-dependent errors are shown in Figs. 6 and 7. Four iterations are required to achieve convergence for the measurements at the latitudes of 9 and 24 deg N, while 3 iterations are sufficient for the other latitudes. Thus, the problems under examination are characterized by moderate degrees of nonlinearity.

The conclusions of our numerical analysis can be summarized as follows:

Table 2
Tangent altitudes for the measurement at the latitude of 79 deg N (state 1)

Number	Tangent altitude (km)
1	46.351
2	43.087
3	39.821
4	36.512
5	33.264
6	29.974
7	26.680
8	23.438
9	20.132

- (1) The agreement between the retrieved profiles is acceptable. The relative error \mathbf{A}_x attains a maximum value of about 7% at the latitude of 39 deg N and at the altitude of 33.25 km. For other latitudes, the relative error does not exceed 5%. Note that the maximum value of the relative error \mathbf{A}_x is accompanied by a large value of the bound of the total error $\mathbf{A}_{\text{total}}$ (between 25 and 40 km). In general, the relation $[\mathbf{A}_x]_i < [\mathbf{A}_{\text{total}}]_i$ is satisfied for all altitudes and latitudes.
- (2) The regularization parameters estimated by the L-curve method are slightly larger than the regularization parameters computed by the a priori minimum bound method. Consequently, the L-curve solutions are characterized by larger smoothing errors and smaller noise errors. The minimum bound solutions are slightly under smoothed and small oscillations are visible at latitudes of 9 deg N and 24 deg N.
- (3) The smoothing error is the dominant error source for both parameter choice methods and this component of the total error increases at low and high altitudes.
- (4) The a priori minimum bound method systematically gives smaller bounds of the total error. Thus, the solutions obtained by using the a priori minimum bound method are closer to the minimum bound solutions than the solutions derived by using the L-curve method.

The applicability of the a priori minimum bound method strongly depends on the data model used. For a data model with

$$\mathbf{y}^\delta = \mathbf{I}_{\text{meas}}^n,$$

$$\mathbf{F} = \mathbf{P}_{\text{sim}} \mathbf{I}_{\text{sim}}^n,$$

where $\mathbf{P}_{\text{sim}} = \mathbf{P}_{\text{sim}}(s, \lambda)$ is a lower-order polynomial whose coefficients are retrieved together with the atmospheric profiles, the degree of nonlinearity is large and the a priori minimum bound method leads to erroneous results.

7. Conclusions

An error analysis for the solution of nonlinear inverse problems solved by Tikhonov regularization has been presented. Our analysis was performed in a semi-stochastic setting and has been applied to inverse problems arising in atmosphere remote sensing. We discussed a diagnostic tool for estimating the components of the error after the iterative process is completed. Essentially, we derived an estimator of the bound of the total error (that possesses a minimum for an optimal value of the regularization parameter) and for its components, the smoothing and the noise errors. An a priori parameter choice method relying on the minimization of an estimator of the bound of the error between the first iterate and the exact solution has been introduced. This parameter choice method was derived under a linearity assumption of the forward model and is applicable for a DOAS type model. Numerical experiments showed that:

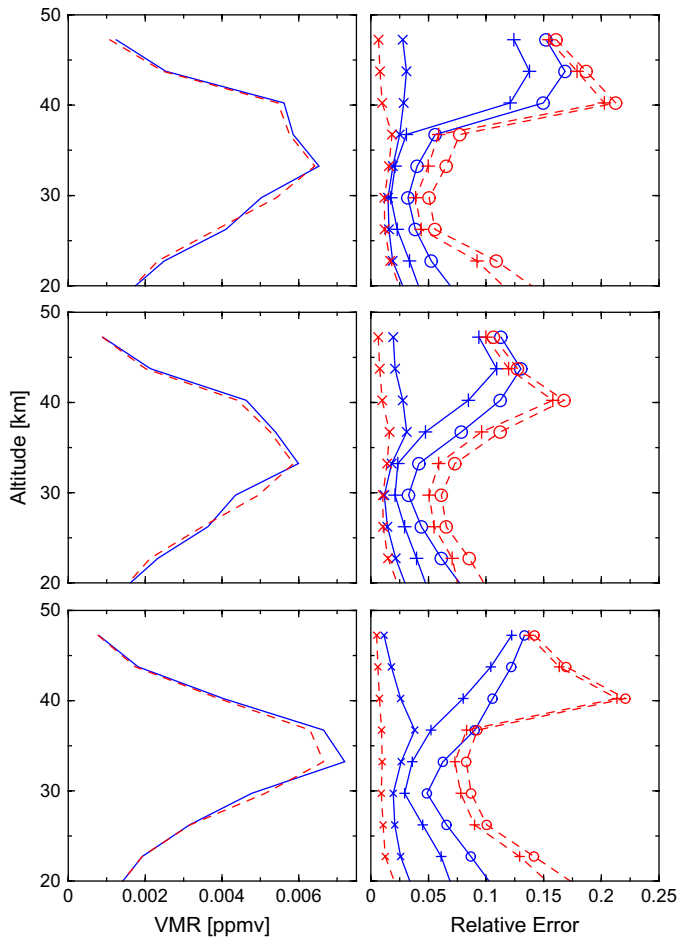


Fig. 6. Retrieved volume mixing ratios (left) and associated errors (right) for SCIAMACHY Orbit 12253, latitudes 9, 24, and 39 deg N. Minimum bound method: solid lines, L-curve method: dashed lines. Smoothing error: +, noise error: x, total error: circle.

- (1) for problems with small degrees of nonlinearity, the value of the regularization parameter estimated by the a priori minimum bound method is close to the value predicted by the L-curve method and to the minimizer of the total error bound estimator, while
- (2) for problems with moderate degrees of nonlinearity, the solutions given by the a priori minimum bound method are characterized by smaller bounds of the total error as compared to those obtained by using the L-curve method.

Appendix A. An estimator for the smoothing error

In this appendix we derive an estimator for the smoothing error vector $\epsilon_{\text{smooth}} = (\mathbf{A}_\lambda - \mathbf{I}_n)(\hat{\mathbf{x}} - \mathbf{x}_a)$. Because the smoothing error does not depend on the measurement error, we consider \mathbf{x}_λ as a minimizer of \mathcal{F} with \mathbf{y}^δ replaced by \mathbf{y} . Clearly, \mathbf{x}_λ satisfies the first order optimality condition

$$\mathbf{K}^T(\mathbf{x}_\lambda)[\mathbf{F}(\mathbf{x}_\lambda) - \mathbf{y}] + \lambda^2 \mathbf{L}^T \mathbf{L}(\mathbf{x}_\lambda - \mathbf{x}_a) = 0. \quad (\text{A.1})$$

Assuming the linearization around \mathbf{x}_λ ,

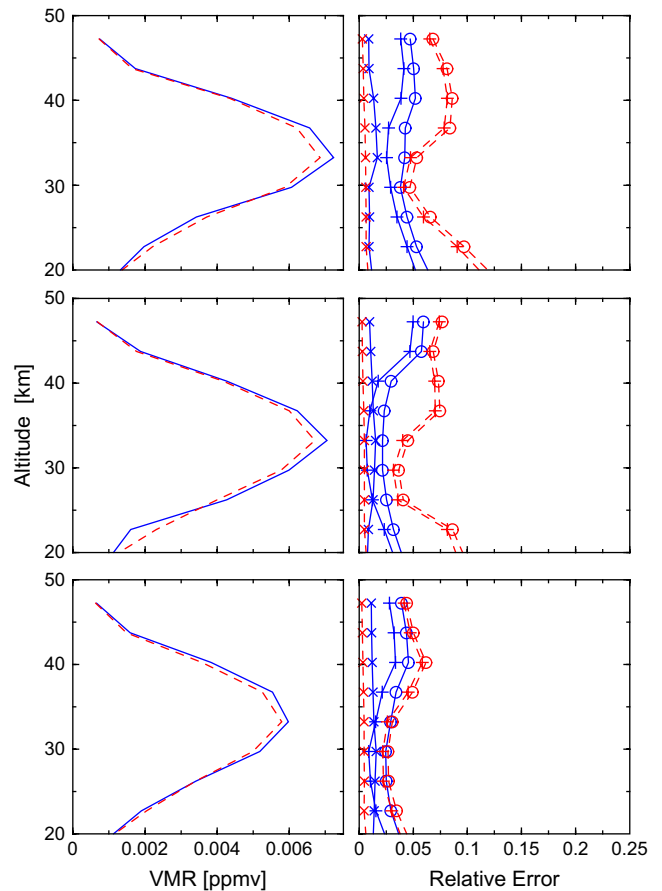


Fig. 7. Same as in Fig. 6 for SCIAMACHY Orbit 12253, latitudes 53, 67, and 79 deg N.

$$\mathbf{y} = \mathbf{F}(\mathbf{x}_\lambda) + \mathbf{K}(\mathbf{x}_\lambda)(\hat{\mathbf{x}} - \mathbf{x}_\lambda) + \mathcal{R}(\mathbf{x}_\lambda, \hat{\mathbf{x}}), \quad (\text{A.2})$$

neglecting $\mathcal{R}(\mathbf{x}_\lambda, \hat{\mathbf{x}})$ and defining the residual for the noise-free case as $\mathbf{r}_\lambda = \mathbf{F}(\mathbf{x}_\lambda) - \mathbf{y}$, we rewrite the above equations as

$$\mathbf{K}^T \mathbf{r}_\lambda + \lambda^2 \mathbf{L}^T \mathbf{L}(\mathbf{x}_\lambda - \mathbf{x}_a) = 0, \quad (\text{A.3})$$

$$\mathbf{r}_\lambda - \mathbf{K}(\mathbf{x}_\lambda - \hat{\mathbf{x}}) = 0. \quad (\text{A.4})$$

Since $\mathbf{x}_\lambda - \mathbf{x}_a = (\mathbf{x}_\lambda - \hat{\mathbf{x}}) + (\hat{\mathbf{x}} - \mathbf{x}_a)$, Eqs. (A.3) and (A.4) can be viewed as a system of matrix equations with the unknown vectors $\mathbf{x}_\lambda - \hat{\mathbf{x}}$ and $\hat{\mathbf{x}} - \mathbf{x}_a$.

Inserting Eq. (A.4) into Eq. (A.3) gives

$$\mathbf{K}^T \mathbf{K}(\mathbf{x}_\lambda - \hat{\mathbf{x}}) + \lambda^2 \mathbf{L}^T \mathbf{L}(\mathbf{x}_\lambda - \mathbf{x}_a) = 0, \quad (\text{A.5})$$

whence

$$\begin{aligned} \mathbf{x}_\lambda - \hat{\mathbf{x}} &= -(\mathbf{K}^T \mathbf{K} + \lambda^2 \mathbf{L}^T \mathbf{L})^{-1} \lambda^2 \mathbf{L}^T \mathbf{L}(\hat{\mathbf{x}} - \mathbf{x}_a) \\ &= (\mathbf{A}_\lambda - \mathbf{I}_n)(\hat{\mathbf{x}} - \mathbf{x}_a), \end{aligned} \quad (\text{A.6})$$

follows. The least squares solution of Eq. (A.4) leads to

$$\mathbf{x}_\lambda - \hat{\mathbf{x}} = (\mathbf{K}^T \mathbf{K})^{-1} \mathbf{K}^T \mathbf{r}_\lambda, \quad (\text{A.7})$$

and we conclude that

$$\epsilon_{\text{smooth}} = (\mathbf{A}_\lambda - \mathbf{I}_n)(\hat{\mathbf{x}} - \mathbf{x}_a) = \mathbf{x}_\lambda - \hat{\mathbf{x}} = (\mathbf{K}^T \mathbf{K})^{-1} \mathbf{K}^T \mathbf{r}_\lambda. \quad (\text{A.8})$$

The above relation shows that ϵ_{smooth} is a deterministic quantity. The residual \mathbf{r}_λ converges to zero as λ tends to zero. This follows from Eq. (A.4) by using the convergence results: $\mathbf{A}_\lambda \rightarrow \mathbf{I}_n$ and $\mathbf{x}_\lambda - \hat{\mathbf{x}} \rightarrow 0$ (c.f. Eq. (A.6)) as $\lambda \rightarrow 0$. Consequently, ϵ_{smooth} converges to zero as λ tends to zero. If we now replace \mathbf{r}_λ by $\mathbf{r}_\lambda^\delta$, where $\mathbf{r}_\lambda^\delta = \mathbf{F}(\mathbf{x}_\lambda^\delta) - \mathbf{y}^\delta$ and assume that $\mathbf{K}(\mathbf{x}_\lambda^\delta) \approx \mathbf{K}(\mathbf{x}_\lambda)$, we arrive at the following implementable expression for the smoothing error vector

$$\epsilon_{\text{smooth}}^\delta = (\mathbf{K}^T \mathbf{K})^{-1} \mathbf{K}^T \mathbf{r}_\lambda^\delta. \quad (\text{A.9})$$

From Eqs. (A.4) and (A.6) we see that

$$\mathbf{r}_\lambda = \mathbf{K}(\mathbf{A}_\lambda - \mathbf{I}_n)(\hat{\mathbf{x}} - \mathbf{x}_a) = (\hat{\mathbf{A}}_\lambda - \mathbf{I}_m) \mathbf{K}(\hat{\mathbf{x}} - \mathbf{x}_a), \quad (\text{A.10})$$

where $\hat{\mathbf{A}}_\lambda$ is the influence matrix, $\hat{\mathbf{A}}_\lambda = \mathbf{K} \mathbf{K}_\lambda^\dagger$, and $\mathbf{K}_\lambda^\dagger$ is given by Eq. (9). Considering the noisy case and the linearization around $\mathbf{x}_\lambda^\delta$ we arrive at

$$\mathbf{r}_\lambda^\delta = (\hat{\mathbf{A}}_\lambda - \mathbf{I}_m) \mathbf{K}(\hat{\mathbf{x}} - \mathbf{x}_a) + (\hat{\mathbf{A}}_\lambda - \mathbf{I}_m) \delta, \quad (\text{A.11})$$

Thus, $\mathbf{r}_\lambda^\delta$ is an unbiased estimator of \mathbf{r}_λ , i.e., $\mathcal{E}\{\mathbf{r}_\lambda^\delta\} = \mathbf{r}_\lambda$, with the covariance matrix

$$\mathbf{S}(\mathbf{r}_\lambda^\delta) = \frac{1}{m} (\hat{\mathbf{A}}_\lambda - \mathbf{I}_m) (\hat{\mathbf{A}}_\lambda - \mathbf{I}_m)^T. \quad (\text{A.12})$$

In contrast to \mathbf{r}_λ , $\mathbf{r}_\lambda^\delta$ does not converge to zero as λ tends to zero and therefore $\epsilon_{\text{smooth}}^\delta$ does not converge to zero as λ tends to zero.

Appendix B. A bound of the noise error and the minimum bound method in a discrete semi-stochastic setting

In this appendix we derive a bound of the noise error in the case $\mathbf{L} = \mathbf{I}_n$ and formulate the minimum bound method in a discrete semi-stochastic setting. We have

$$\|\epsilon_{\text{noise}}^\delta\|^2 = \delta^T \mathbf{K}_\lambda^{\dagger T} \mathbf{K}_\lambda^\dagger \delta = \delta^T \mathbf{K}_\lambda^{\dagger T} (\mathbf{K}^T \mathbf{K} + \lambda^2 \mathbf{L}^T \mathbf{L})^{-1} \mathbf{K}^T \delta \quad (\text{B.1})$$

The identity

$$(\mathbf{K}^T \mathbf{K} + \lambda^2 \mathbf{L}^T \mathbf{L})^{-1} \mathbf{K}^T = \frac{1}{\lambda^2} (\mathbf{L}^T \mathbf{L})^{-1} \mathbf{K}^T \mathbf{A}_{1\lambda}, \quad (\text{B.2})$$

with

$$\mathbf{A}_{1\lambda} = \left[\frac{1}{\lambda^2} \mathbf{K} (\mathbf{L}^T \mathbf{L})^{-1} \mathbf{K}^T + \mathbf{I}_m \right]^{-1}, \quad (\text{B.3})$$

yields, for $\mathbf{L} = \mathbf{I}_n$,

$$\|\epsilon_{\text{noise}}^\delta\|^2 = \frac{1}{\lambda^2} \delta^T \mathbf{K}_\lambda^{\dagger T} \mathbf{K}^T \mathbf{A}_{1\lambda} \delta \leq \frac{1}{\lambda^2} \|\mathbf{K} \mathbf{K}_\lambda^\dagger\| \|\mathbf{A}_{1\lambda} \delta\|. \quad (\text{B.4})$$

Relying on the SVD of the matrix \mathbf{K} , it is easy to show that

$$\frac{1}{\lambda^2} \|\mathbf{A}_{1\lambda} \delta\| \leq \frac{1}{\lambda^2} \|\delta\|, \quad (\text{B.5})$$

$$\|\mathbf{K} \mathbf{K}_\lambda^\dagger \delta\| \leq C \|\delta\|, \quad (\text{B.6})$$

where C is a positive constant such that

$$\frac{\gamma_i^2}{\gamma_i^2 + \lambda^2} \leq C \leq 1, \quad (\text{B.7})$$

for all singular values γ_i , $i = 1, \dots, n$, of the matrix \mathbf{K} . Thus,

$$\mathcal{E}\{\|\epsilon_{\text{noise}}^\delta\|^2\} \leq \frac{C}{\lambda^2} \mathcal{E}\{\|\delta\|^2\} = \frac{C}{\lambda^2}. \quad (\text{B.8})$$

In this case, taking into account Eqs. (14) and (A.8), the bound of the total error may be written as

$$Y(\lambda^2) = 2 \left(\|\mathbf{x}_\lambda - \hat{\mathbf{x}}\|^2 + \frac{C}{\lambda^2} \right). \quad (\text{B.9})$$

The above relation is the starting point for the minimum bound method. Minimization of the function

$$\frac{1}{2} \|\mathbf{x}_\lambda - \hat{\mathbf{x}}\|^2 + \frac{C}{2\lambda^2}, \quad (\text{B.10})$$

with respect to λ^2 gives an a posteriori strategy for computing the optimal regularization parameter. To derive the necessary condition for such a minimum, we consider the function

$$f(\lambda^2) = \frac{1}{2} \|\mathbf{x}_\lambda - \hat{\mathbf{x}}\|^2 \quad (\text{B.11})$$

and obtain

$$\frac{df}{d\lambda^2}(\lambda^2) = (\mathbf{x}_\lambda - \hat{\mathbf{x}})^T \frac{d\mathbf{x}_\lambda}{d\lambda^2}. \quad (\text{B.12})$$

Formal differentiation of Eq. (A.1) with respect to λ^2 yields

$$\frac{d\mathbf{K}^T}{d\lambda^2} [\mathbf{F}(\mathbf{x}_\lambda) - \mathbf{y}] + \mathbf{K}^T \mathbf{K} \frac{d\mathbf{x}_\lambda}{d\lambda^2} + (\mathbf{x}_\lambda - \mathbf{x}_a) + \lambda^2 \frac{d\mathbf{x}_\lambda}{d\lambda^2} = 0$$

whence, neglecting the first term, gives

$$\frac{d\mathbf{x}_\lambda}{d\lambda^2} = (\mathbf{K}^T \mathbf{K} + \lambda^2 \mathbf{I}_n)^{-1} (\mathbf{x}_a - \mathbf{x}_\lambda). \quad (\text{B.13})$$

Further, using the relation $\mathbf{x}_a - \mathbf{x}_\lambda = \mathbf{K}^T \mathbf{r}_\lambda / \lambda^2$ (Eq. (A.3)), the identity (B.2) and the linearity relation $(\mathbf{x}_\lambda - \hat{\mathbf{x}})^T \mathbf{K}^T = \mathbf{r}_\lambda^T$ (Eq. (A.4)), we arrive at

$$\frac{df}{d\lambda^2}(\lambda^2) = \frac{1}{\lambda^2} \mathbf{r}_\lambda^T (\mathbf{K} \mathbf{K}^T + \lambda^2 \mathbf{I}_m)^{-1} \mathbf{r}_\lambda. \quad (\text{B.14})$$

Replacing \mathbf{r}_λ by $\mathbf{r}_\lambda^\delta$ we obtain an implementable parameter choice rule to compute the optimal regularization parameter as the solution of the equation

$$\lambda^2 \mathbf{r}_\lambda^{\delta T} (\mathbf{K} \mathbf{K}^T + \lambda^2 \mathbf{I}_m)^{-1} \mathbf{r}_\lambda^\delta = \frac{C}{2}. \quad (\text{B.15})$$

References

- Amato, U., Masiello, G., Serio, C., Viggiano, M., 2002. The σ -IASI code for the calculation of infrared atmospheric radiance and its derivatives. *Environ. Model. Softw.* 17, 651–667, ERRATUM: vol. 18, p. 97, 2003.
- Bovensmann, H., Burrows, J., Buchwitz, M., Frerick, J., Noël, S., Rozanov, V., Chance, K., Goede, A., 1999. SCIAMACHY: mission objectives and measurement mode. *J. Atmos. Sci.* 56, 127–150.
- Box, M., 2002. Radiative perturbation theory: a review. *Environ. Model. Softw.* 17, 95–106.
- Carissimo, A., Feis, I.D., Serio, C., 2005. The physical retrieval methodology for IASI: the δ -IASI code. *Environ. Model. Softw.* 20, 1111–1126.
- Doicu, A., Schreier, F., Hess, M., 2002. Iteratively regularized Gauss–Newton method for atmospheric remote sensing. *Comput. Phys. Commun.* 148, 214–226.
- Doicu, A., Schreier, F., Hess, M., 2003. Iteratively regularized Gauss–Newton method for bound-constraint problems in atmospheric remote sensing. *Comput. Phys. Commun.* 153 (1), 59–65.
- Doicu, A., Schreier, F., Hess, M., 2005. An iterative regularization method with B-spline approximation for atmospheric temperature and concentration retrievals. *Environ. Model. Softw.* 20, 1101–1109.
- Engl, H., Hanke, M., Neubauer, A., 1996. *Regularization of Inverse Problems*. Kluwer, Dordrecht, NL.
- Eriksson, J., 1996. *Optimization and regularization of nonlinear least squares problems*. Ph.D. thesis, Department of Computing Science, Umea University, Sweden.
- Flittner, D., Bhartia, P., Herman, B., 2000. O₃ profiles retrieved from limb scatter measurements: theory. *Geophys. Res. Lett.* 27, 2601–2604.
- Hansen, P., 1998. *Rank-deficient and Discrete Ill-posed Problems: Numerical Aspects of Linear Inversion*. SIAM, Philadelphia, PA.
- Lukas, M., 1998. Asymptotic behavior of the minimum bound method for choosing the regularization parameter. *Inverse Probl.* 14, 149–159.
- McLinden, C., McConnell, J., Griffioen, E., McElroy, C., 2002. A vector radiative-transfer model for the Odin/OSIRIS project. *Can. J. Phys.* 80, 375–393. DOI: 10.1139/P01-156.
- Neubauer, A., 1989. Tikhonov regularisation for non-linear ill-posed problems: optimal convergence rates and finite-dimensional approximation. *Inverse Probl.* 5, 541–557.
- O’Sullivan, F., Wahba, G., 1985. A cross validated Bayesian retrieval algorithm for nonlinear remote sensing experiments. *J. Comput. Phys.* 59, 441–455.
- Platt, U., 1994. Differential optical absorption spectroscopy (DOAS). In: Sigrist, M. (Ed.), *Air Monitoring by Spectroscopic Techniques*. J. Wiley, pp. 27–84.
- Raus, T., 1984. On the discrepancy principle for the solution of ill-posed problems. *Acta et Commentationes Universitatis Tartuensis de Mathematica* 672, 16–26 (in Russian).
- Rodgers, C., 2000. *Inverse Methods for Atmospheric Sounding: Theory and Practise*. World Scientific, Singapore.
- Schreier, F., Böttger, U., 2003. MIRART, a line-by-line code for infrared atmospheric radiation computations incl. derivatives. *Atmos. Ocean. Opt.* 16, 262–268.
- Spurr, R., Kurosu, T., Chance, K., 2000. A linearized discrete ordinate radiative transfer model for atmospheric remote-sensing retrieval. *J. Quant. Spectrosc. Radiat. Transf.* 68 (6), 689–735.
- Steck, T., 2002. Methods for determining regularization for atmospheric retrieval problems. *Appl. Opt.* 41, 1788–1797.
- Tautenhahn, U., 1994. Error estimates for regularized solutions of nonlinear ill-posed problems. *Inverse Probl.* 10, 485–500.
- Tautenhahn, U., 1997. On a general regularization scheme for nonlinear ill-posed problems. *Inverse Probl.* 13, 1427–1437.